Indicator Principal Component Kriging

V. Suro-Pérez and A. G. Journel

An alternative to multiple indicator kriging is proposed which approximates the full coindicator kriging system by kriging the principal components of the original indicator variables. This transformation is studied in detail for the biGaussian model. It is shown that the cross-correlations between principal components are either insignificant or exactly zero. This result allows derivation of the conditional cumulative density function (cdf) by kriging principal components and then applying a linear back transform. A performance comparison based on a real data set (Walker Lake) is presented which suggests that the proposed method achieves approximation of the conditional cdf equivalent to indicator cokriging but with substantially less variogram modeling effort and at smaller computational cost.

KEY WORDS: indicator kriging, principal component analysis, biGaussian model, indicator covariance matrix, orthogonalization.

INTRODUCTION

Most Earth Sciences data feature patterns of spatial continuity which can be used to model and assess the uncertainty prevailing at unsampled locations. Models of spatial continuity allow going beyond the actual data toward an assessment of the uncertainty specific to each unsampled location. Characterization of uncertainty and spatial interpolation are a primary goal of any geostatistical approach.

The uncertainty associated with an unsampled value \( z(x) \) at location \( x \) can be modeled by the probability distribution of a random function (RF) \( Z(x) \). This distribution is made conditional on the surrounding information. The degree to which the distribution of \( Z(x) \) is influenced by the surrounding data is dictated by the prior model of spatial continuity or dependence between the \((n + 1)\) RVs \( Z(x), Z(x_\alpha), \alpha = 1, \ldots, n \). Consider then the conditional cumulative distribution function (cdf) of \( Z(x) \):

\[
F(x; z \mid \{n\}) = P\{Z(x) \leq z \mid \{Z(x_\alpha) = z(x_\alpha), \alpha = 1, \ldots n\}\}
\]

1Manuscript received 17 July 1990; accepted 11 January 1991.
2Applied Earth Sciences Department, Stanford University, Stanford, CA 94305-2225.
For any set \( \{n\} \) of data values, there are as many conditional cdf's of type \( F(x; z | \{n\}) \) as there are models of spatial continuity which relate the unknown value \( z(x) \) to the data values. All such conditional cdf's \( F_l(x; z | \{n\}) \), \( l = 1, \ldots, L \), including those provided by all forms of indicator kriging, should be seen as alternative models of the uncertainty prevailing at the unsampled location \( x \), rather than different estimates of an elusive \textit{"true"} cdf \( F(x; z | \{n\}) \).

Prior to selection of any estimated value \( z^*(x) \), access to a conditional cdf model \( F_l(x; z | \{n\}) \) allows determination of probability intervals and probabilities of exceedance such as:

\[
P\{Z(x) \in (z_1, z_2) | \{n\}\} = F_l(x; z_2 | \{n\}) - F_l(x; z_1 | \{n\})
\]

\[
P\{Z(x) > z_1 | \{n\}\} = 1 - F_l(x; z_1 | \{n\})
\]

The concept of a loss function (Journel, 1984) allows deriving from the model \( F_l(x; z | \{n\}) \) not just one but as many optimal estimates for \( z(x) \) as there are different criteria for optimality. This usually involves the minimization of the expected value of some loss function associated with the estimation error \( z^*(x) - z(x) \).

Disjunctive Kriging (DK) (Matheron, 1976; Rivoirard, 1989), Indicator Kriging (IK) (Journel, 1983), Multigaussian Kriging (MG) (Verly, 1983), Probability Kriging (PK) (Sullivan, 1984), Uniform Conditioning (Guibal and Remacre, 1984), and BiGaussian Kriging (Marcotte and David, 1985) are all techniques providing models for such conditional cdf's. Except for IK and PK, all these algorithms are parametric in the sense that they call for a prior bivariate or multivariate distribution model for the random function \( Z(x) \); the parameters of the conditional cdf \( F_l(x; z | \{n\}) \) are then derived by some form of kriging. In the case of IK and PK, the bivariate distribution of any pair of RFs \( Z(x) \) and \( Z(x') \) is inferred directly from data rather than derived from prior models.

Principal Component Analysis (PCA) (Anderson, 1984; Borgman and Frahm, 1976; Davis and Greenes, 1983) and the indicator formalism where any attribute value \( z(x) \) is coded into an indicator vector of 0's and 1's, are used in this paper to model the conditional cdf. The approach presented here attempts to fill the gap between the IK approach which uses indicator autocovariances to model \( F_l(x; z | \{n\}) \) and the theoretically better Indicator Cokriging (CoIK) which makes use of both indicator autocovariances and cross-covariances. The proposed approach provides a model of the conditional cdf by kriging the indicator principal components instead of the indicators themselves.

Presentation of the principal components approach is accomplished first by considering a parametric point of view. The biGaussian model is studied in detail and analytical expressions for the corresponding indicator crosscovariances are derived. It is shown that these indicator crosscovariances are not negligible with respect to indicator autocovariances. CoIK would account for such